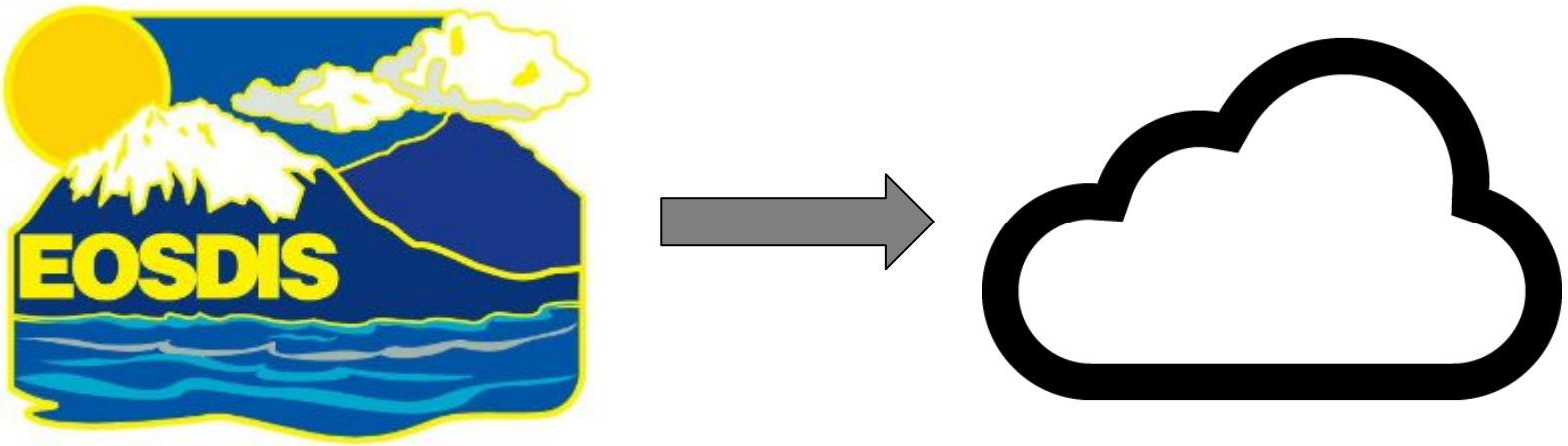
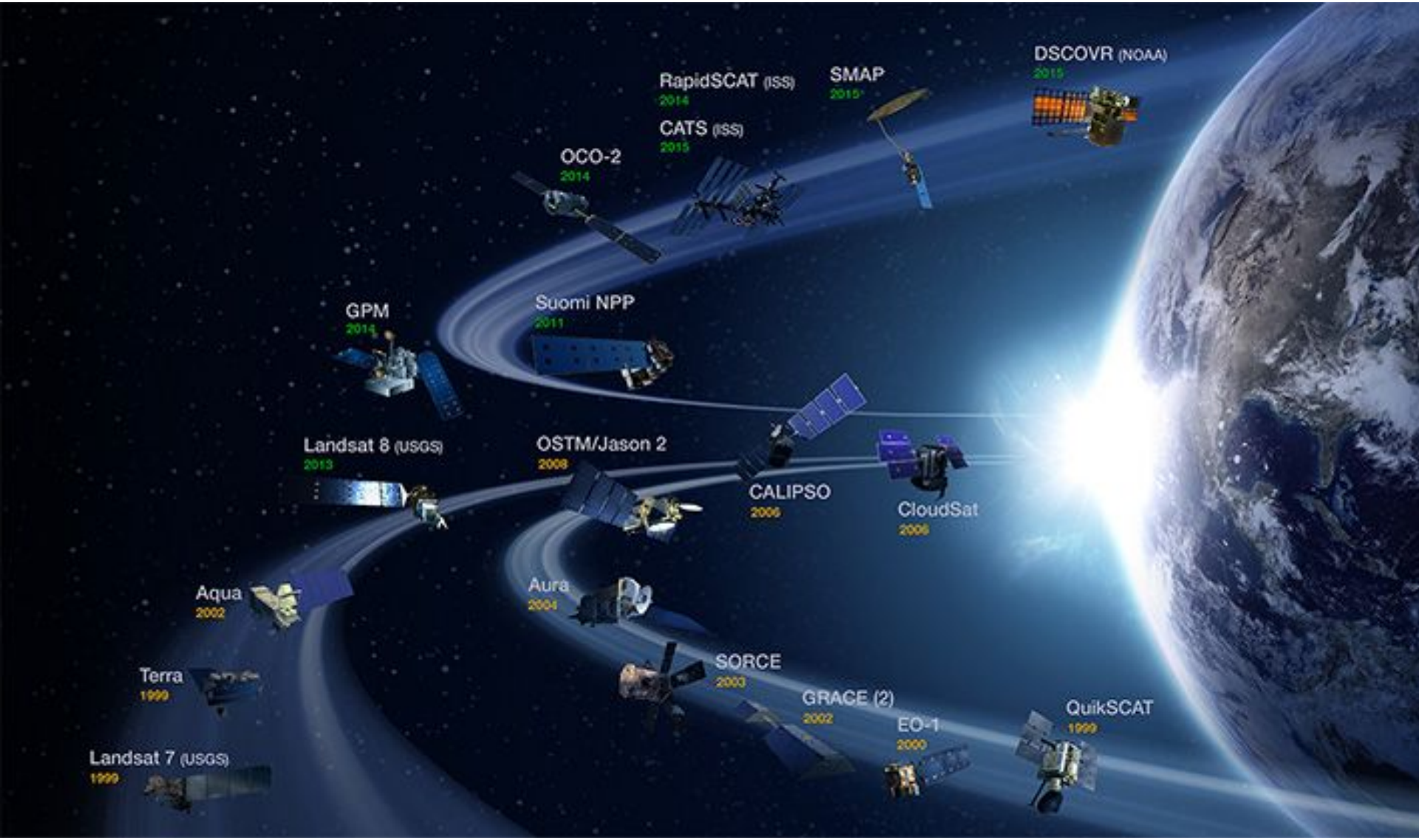
Migrating the Earth Observing System Data and Information System into the Cloud

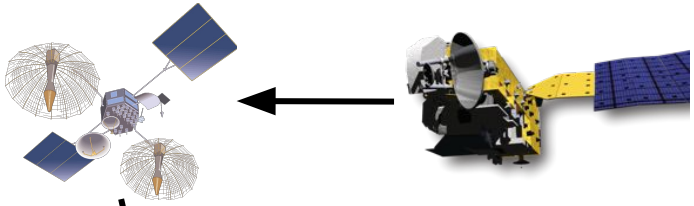


Chris Lynnes
with contributions from Katie Baynes
NASA/GSFC
Earth Observing System Data and Information System

EOSDIS processes, archives and distributes data from Earth observing satellites



EOSDIS Functions



data downlink

EOSDIS

capture and clean

process

archive

subset

distribute

Research

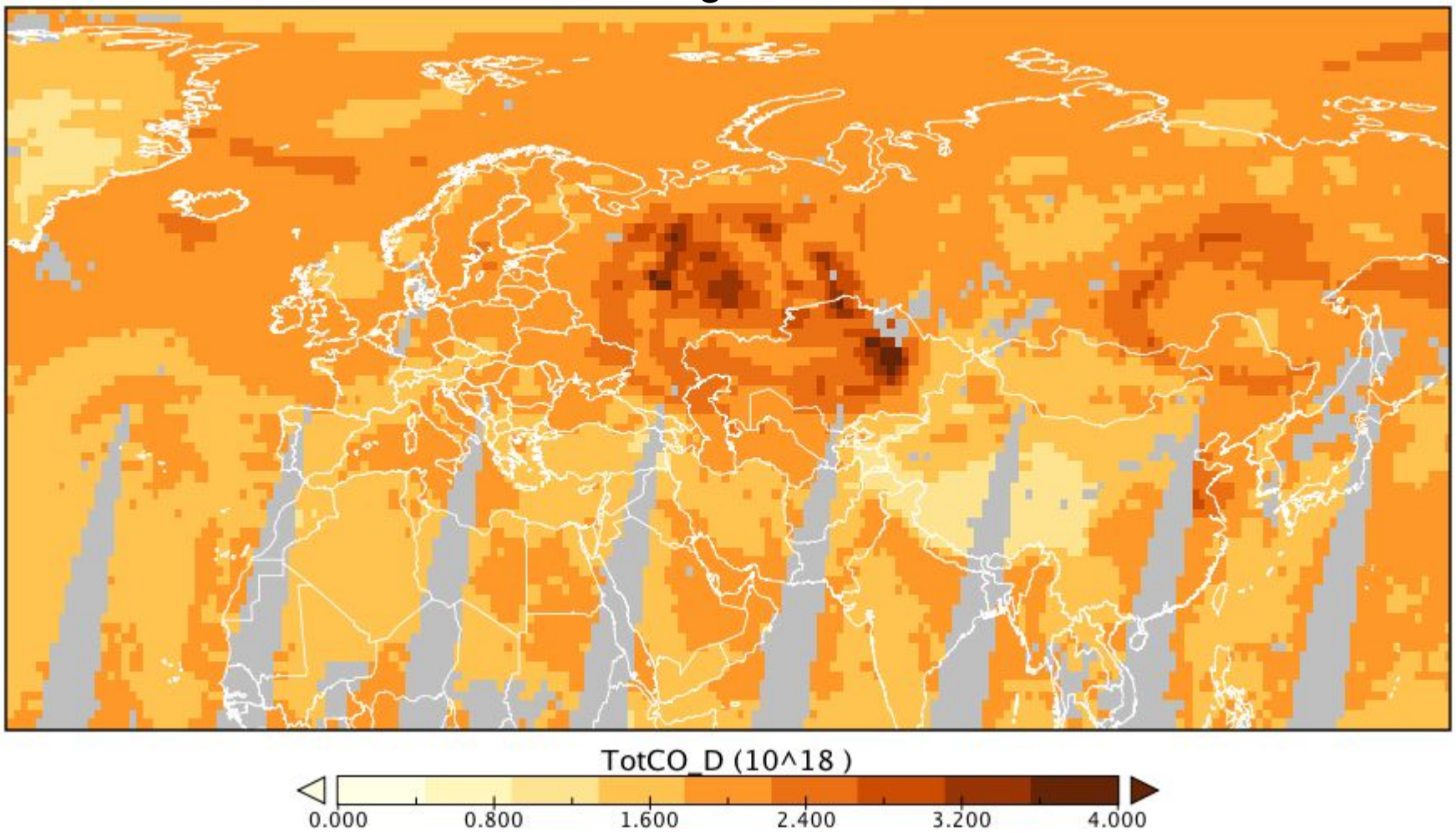
Applications

Education

Users

Example: Atmospheric Infrared Sounder (Aqua Satellite)

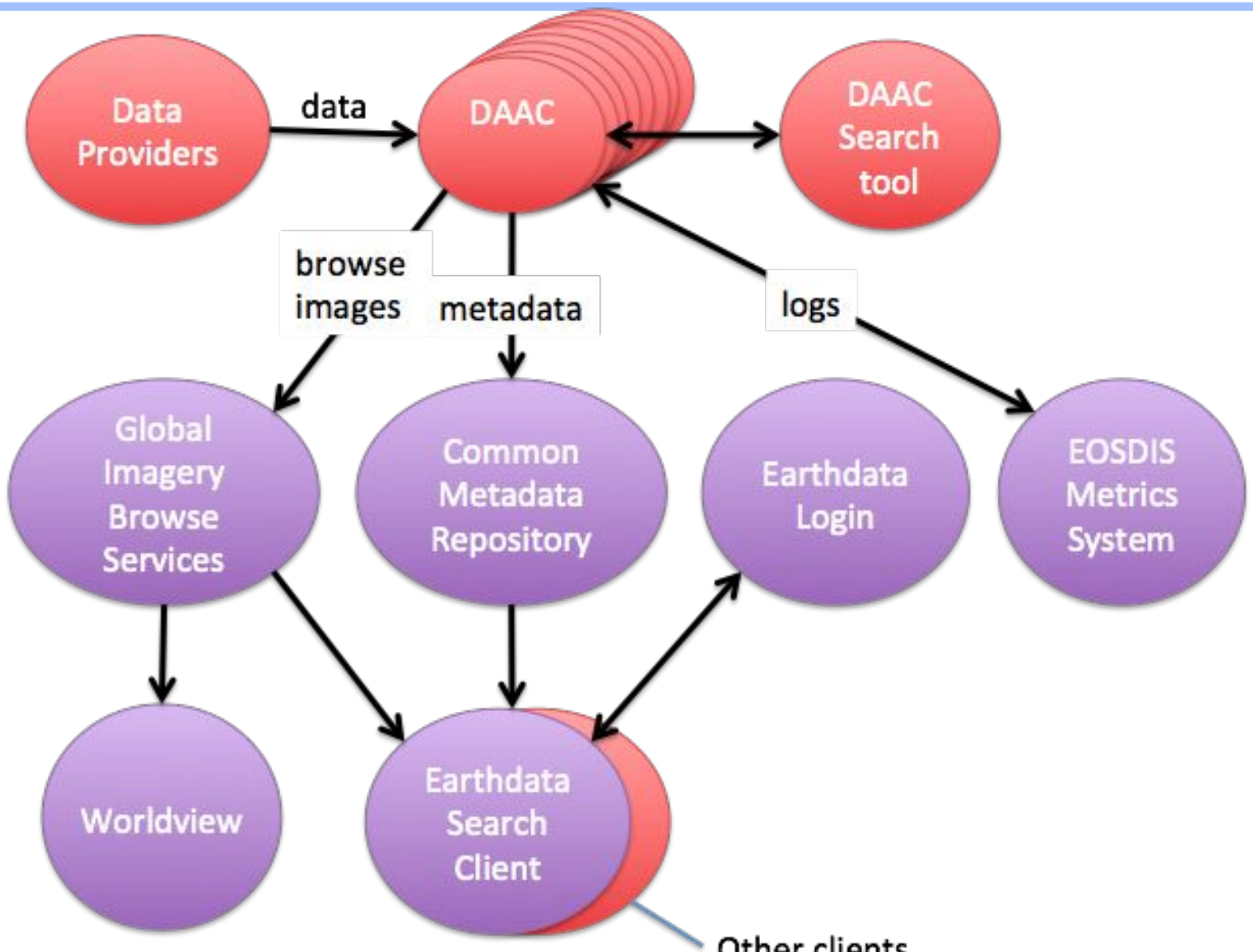
Total Column Carbon Monoxide, Night
11 Aug 2011



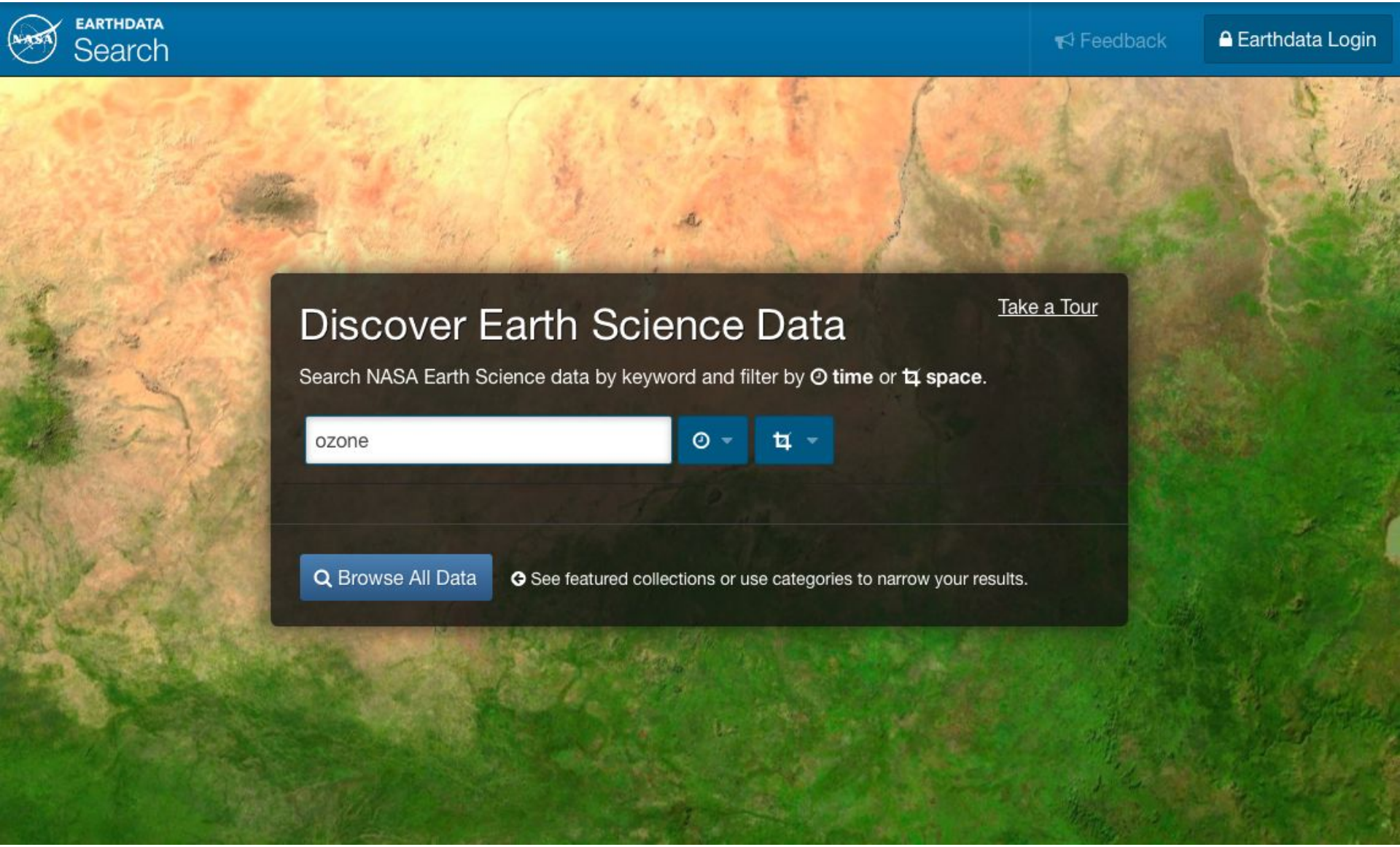
http://earthdata.nasa.gov

EOSDIS Comprises Discipline-Focused Distributed Active Archive Centers

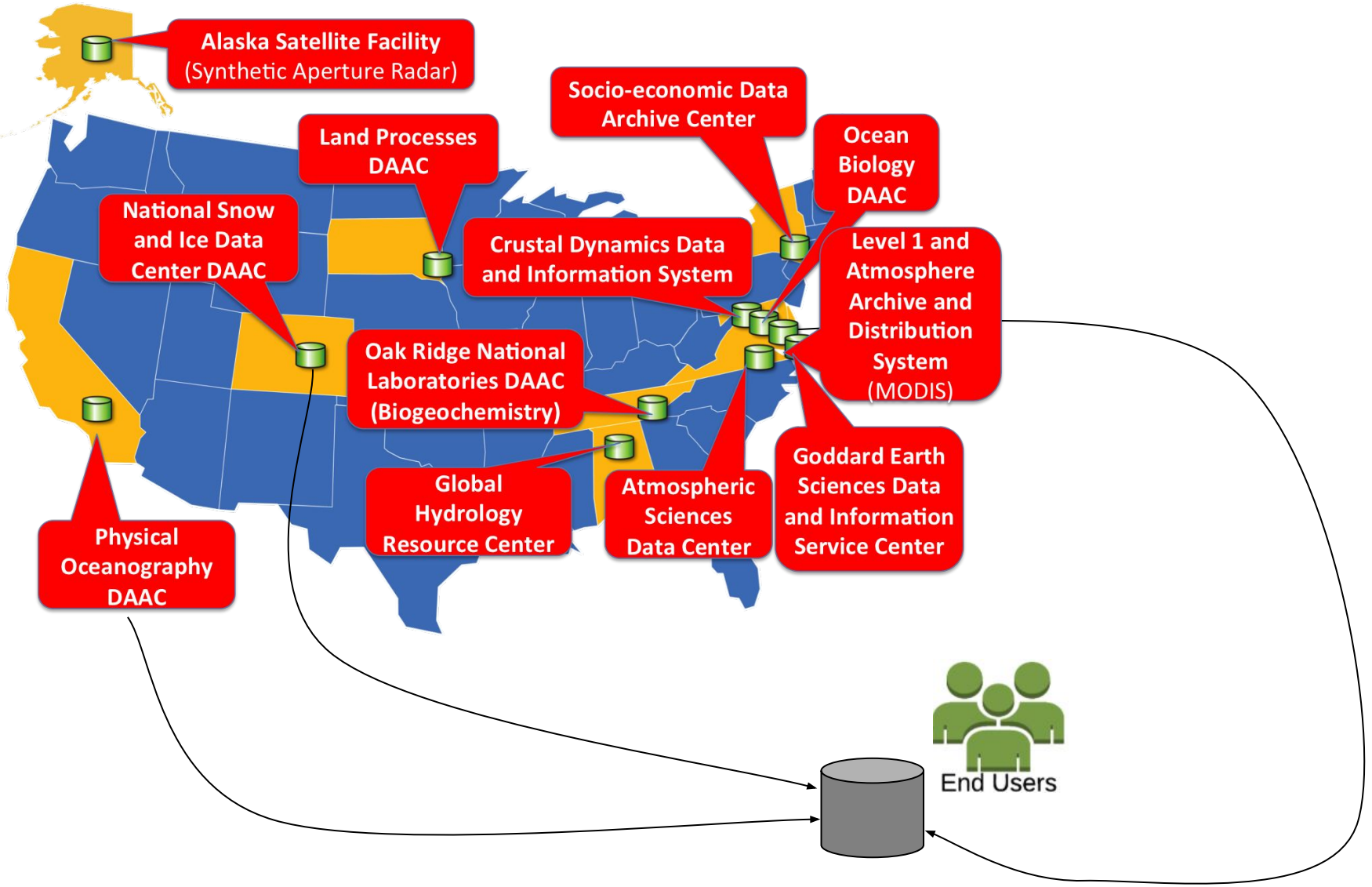
DAACs and users are supported by EOSDIS Common Services



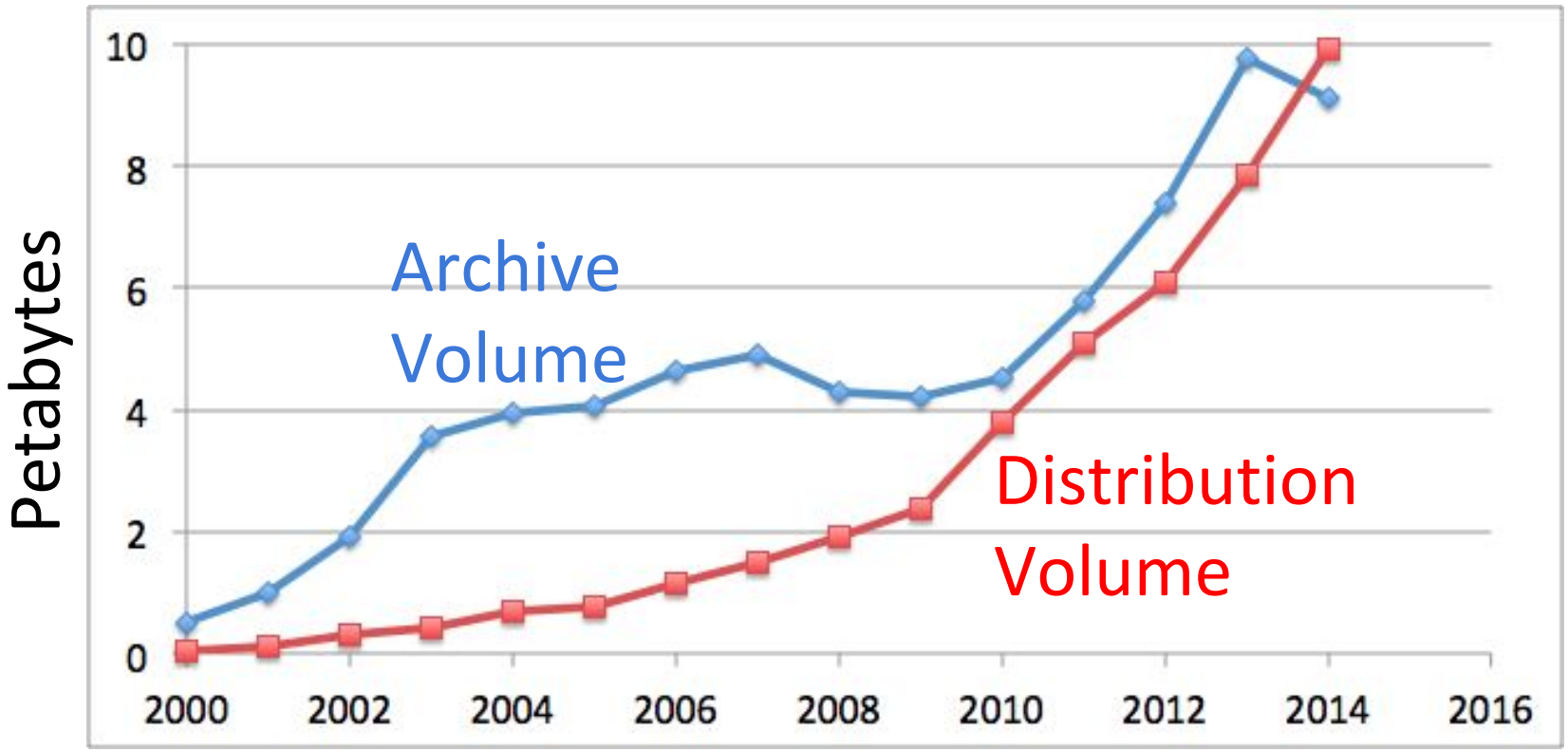
Earthdata Search Client provides one-stop shopping



BUT, users currently need to download data to work on them



EOSDIS Works with (pretty) Big Data



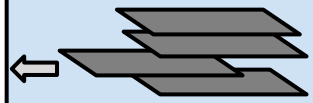
EOSDIS in the Big Data epoch will *enable* more analysis closer to the data.

“More Analysis”

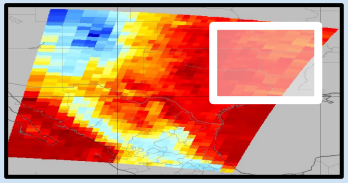
More Complexity

Subset

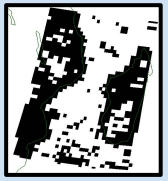
Data Variable



Spatial Area

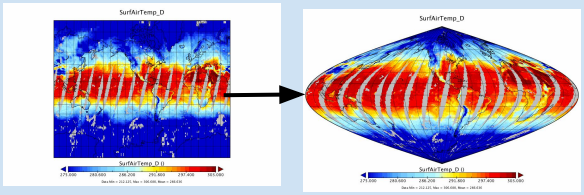


Quality Filter

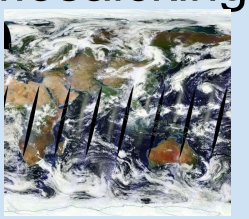


Transform

Reprojection

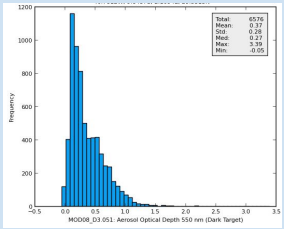


Mosaicking

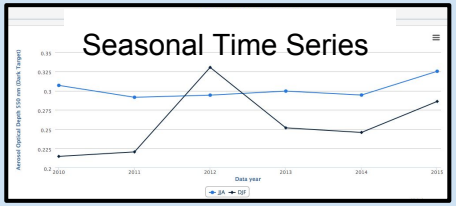


Analyze

Simple Stats



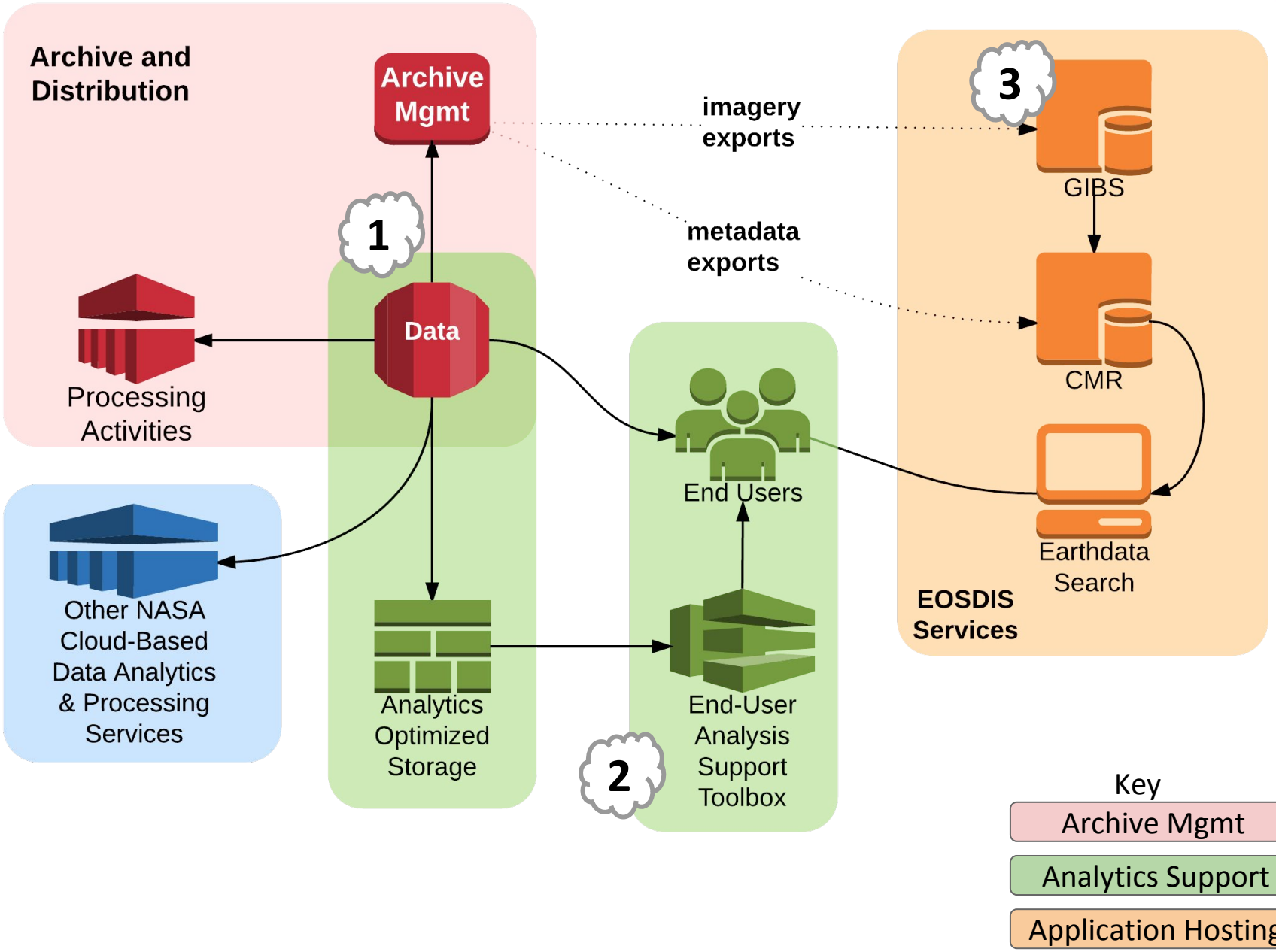
Complex Stats



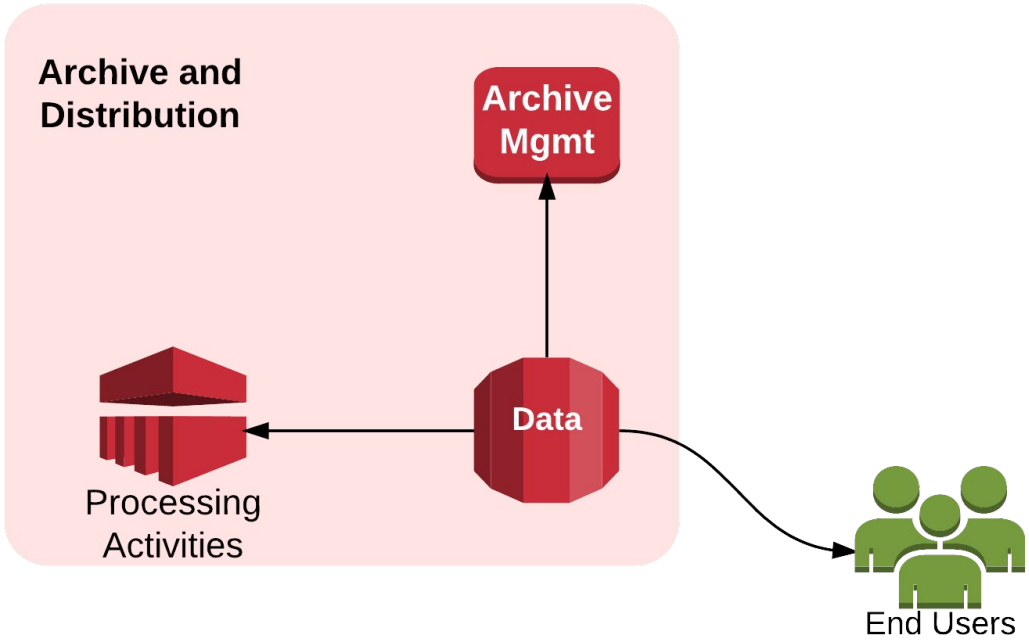
End User’s Algorithm

EOSDIS Cloud Plans

Cloud Prototypes



1 Archive and Distribution Prototype(s)



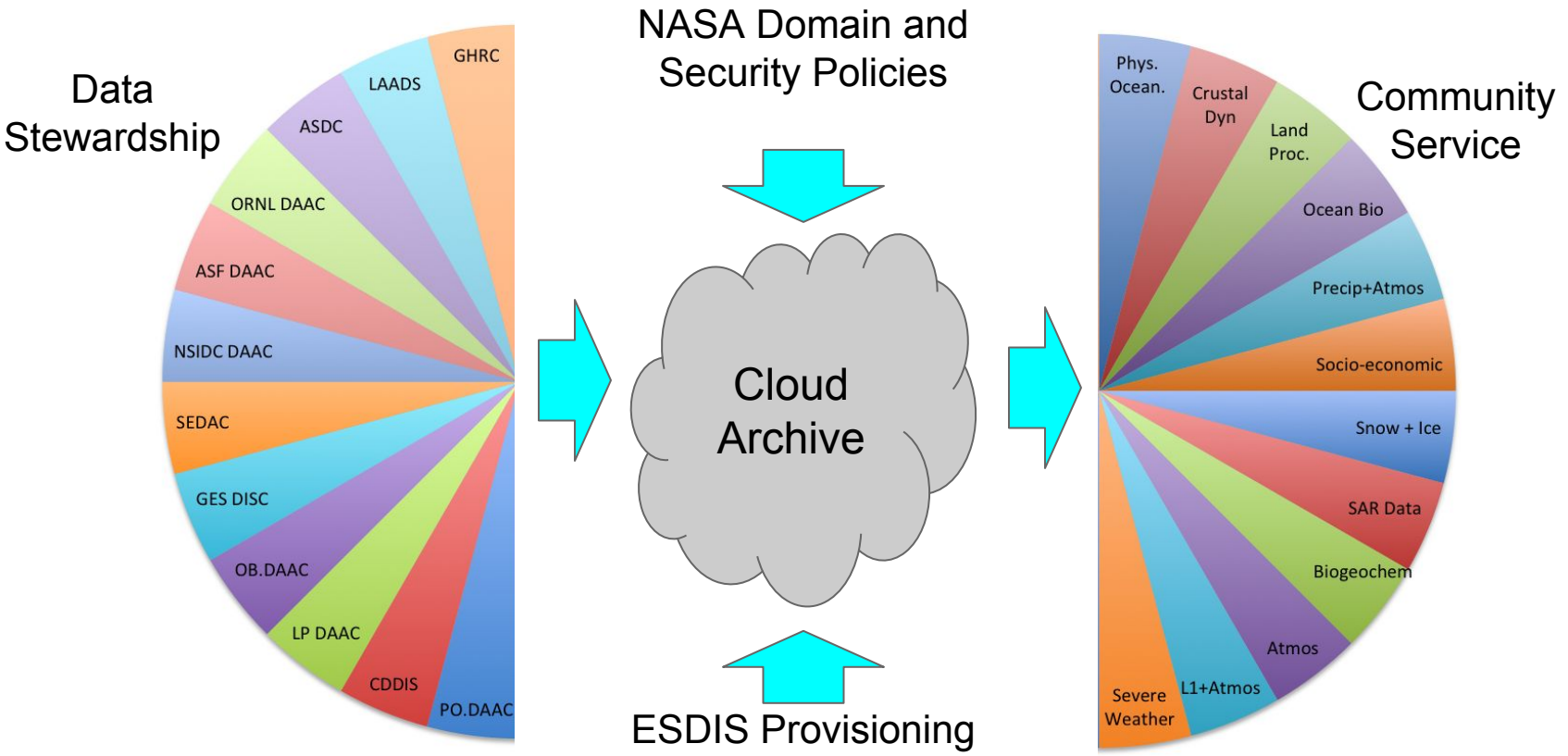
Cloud-based data distribution of Sentinel radar data by the Alaska Satellite Facility DAAC

1 Archive and Distribution in the Cloud

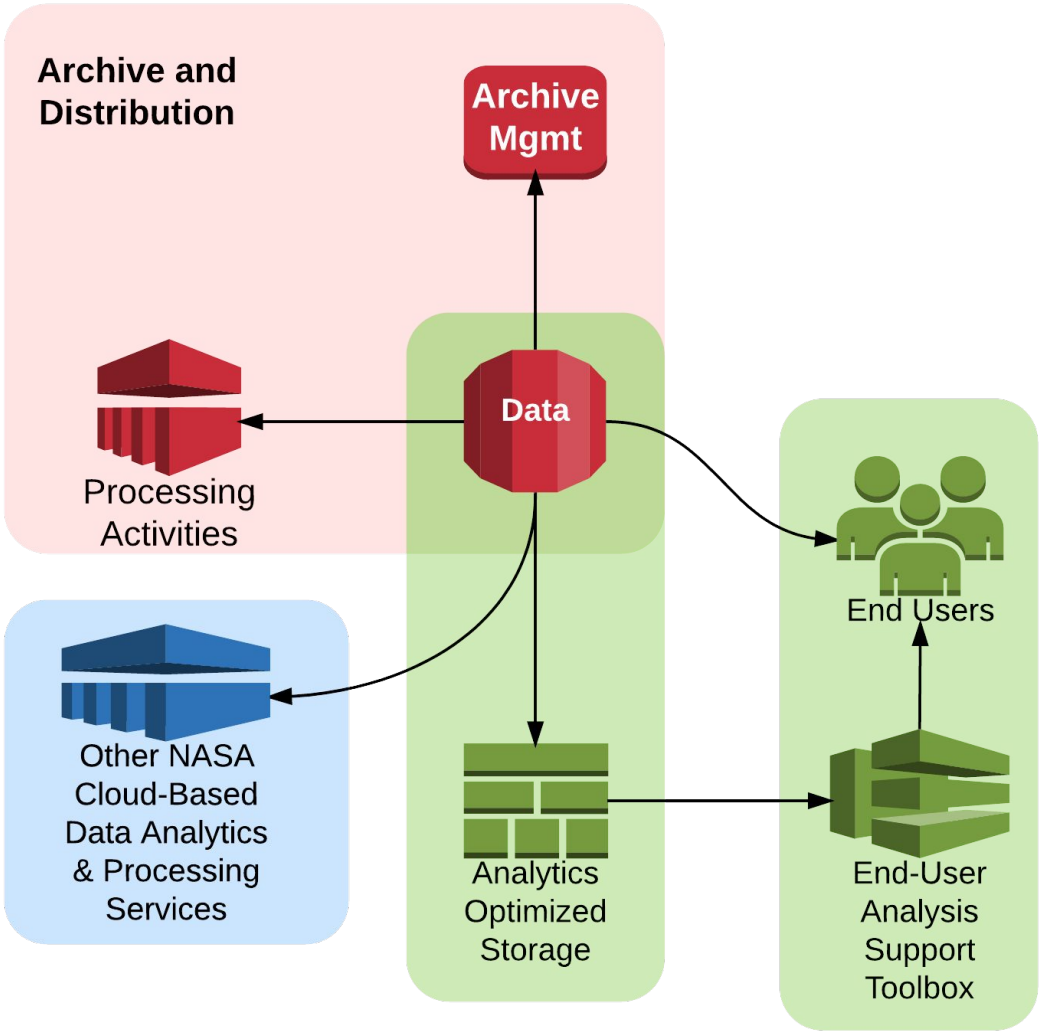
- Science User Benefit: Use bigger datasets
- Cost Savings
  - *Possible* savings for Program
  - Definite savings to *facilities*
- **BUT**:
  - Must solve the egress charge policy for the average scientist
  - Requires user paradigm shift

1 Data Management Paradigm Shift

- ESDIS buys & provisions storage
- 1 DAAC provides archiving services for a data product
- Any DAAC can add services to any product to serve their user community



2 Cloud Analytics Prototypes



Analysis Support Toolbox to attract users to cloud analytics

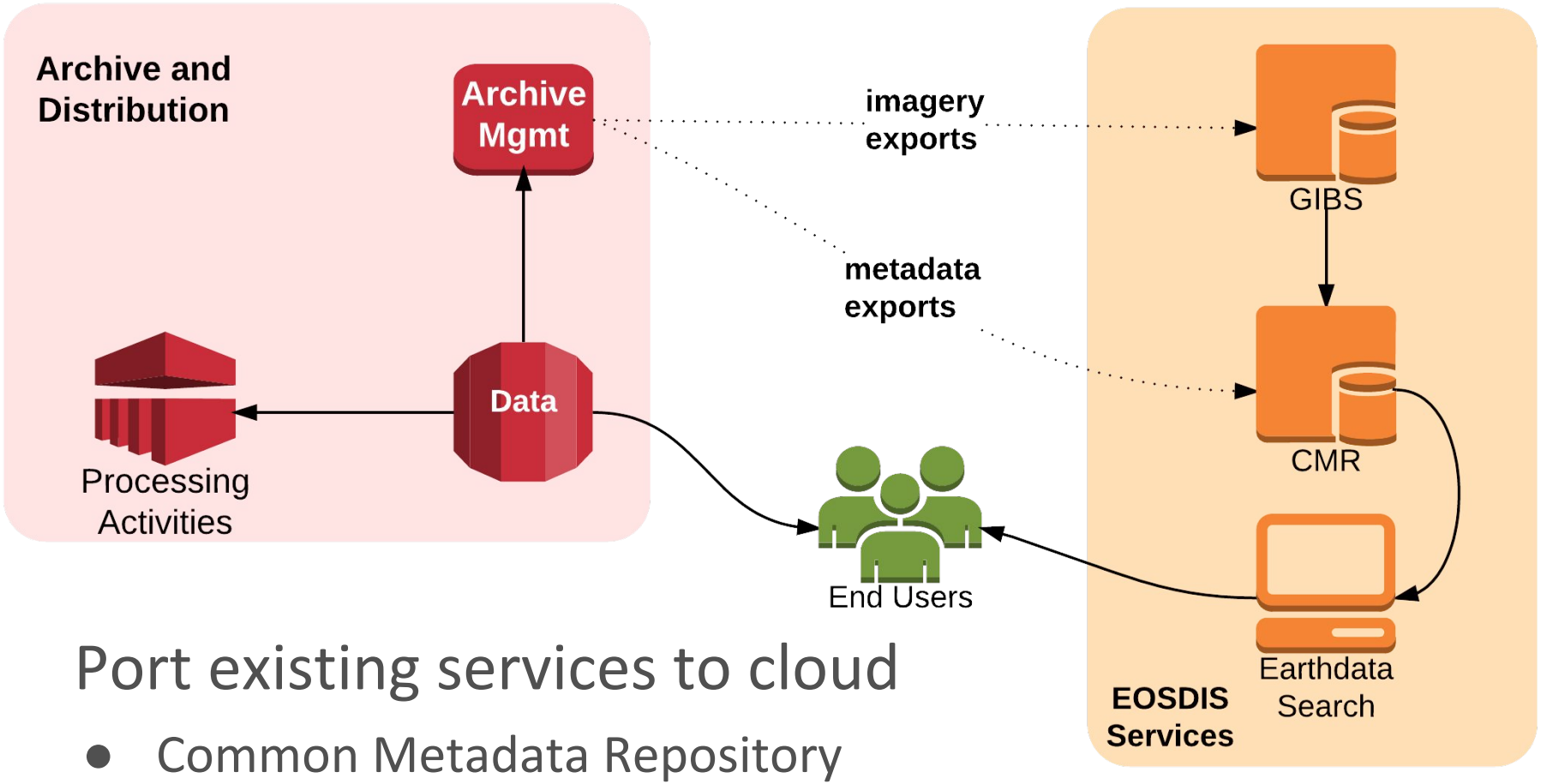
- Community open source tools
- DAAC-developed tools
- Cloud analytics examples and recipes

Analytics Optimized Storage for Big Data Analytics

2 Analytics in the Cloud

- Long Term Paradigm Shift
  - Scientists work on data “in place”
  - High-value data are (also) in databases
- Science User Benefits
  - Analyze very large datasets
    - Any subset slice of data
    - Long time series
  - Avoid data management drudgery
  - Reuse code from colleagues

3 Application-Hosting Prototypes

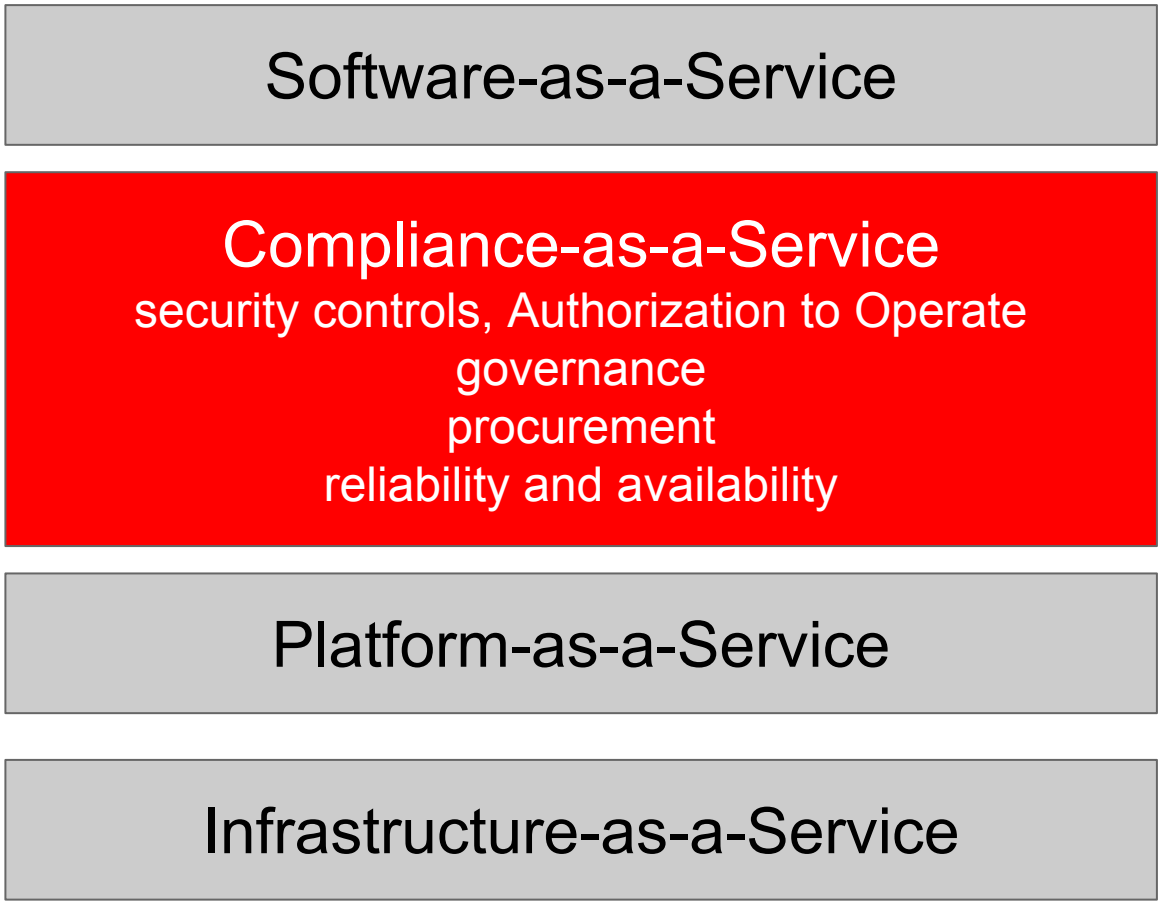


Port existing services to cloud

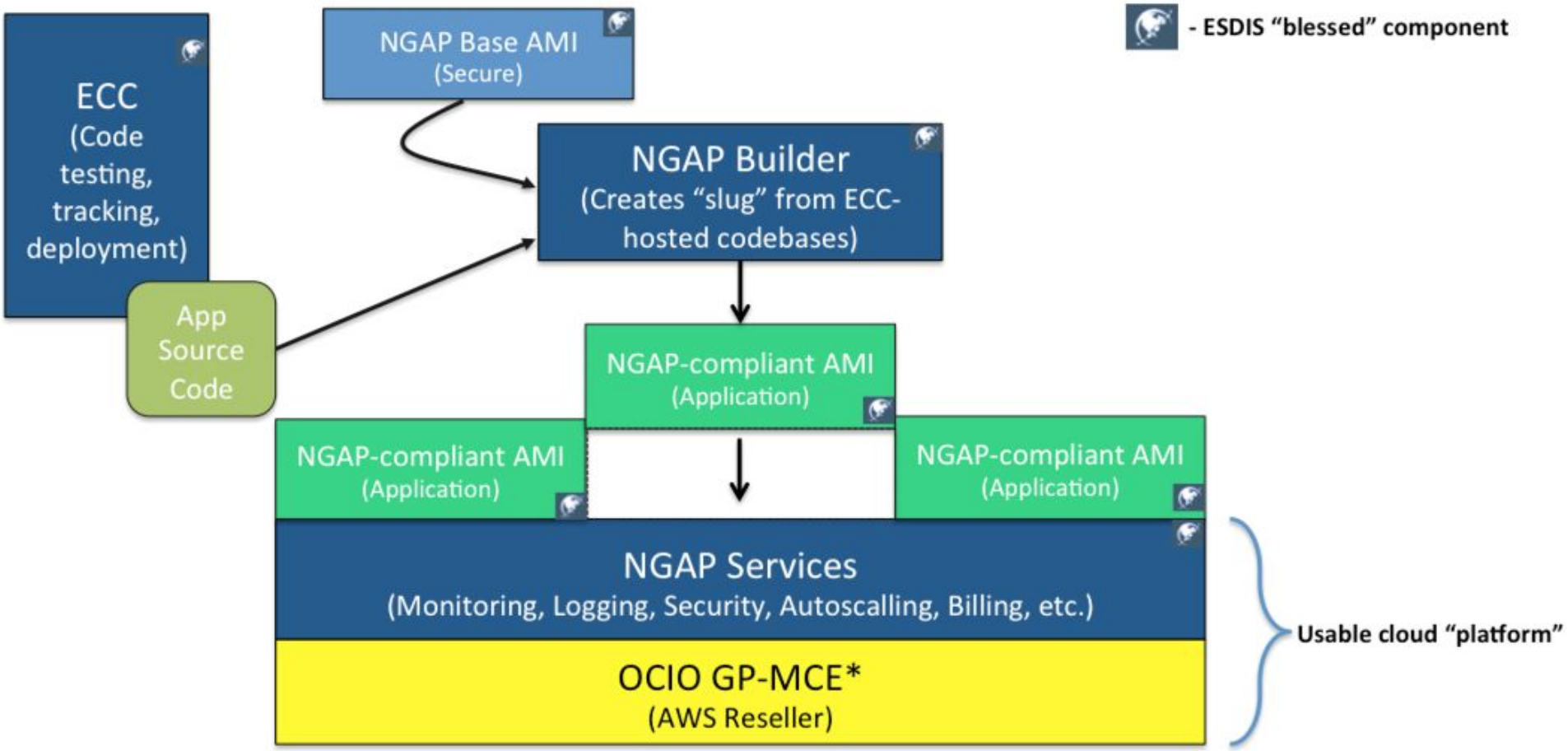
- Common Metadata Repository
- Global Imagery Browse System
- Earthdata Search client

Leveraging a reusable EOSDIS framework NGAP...

“Compliance-as-a-Service”



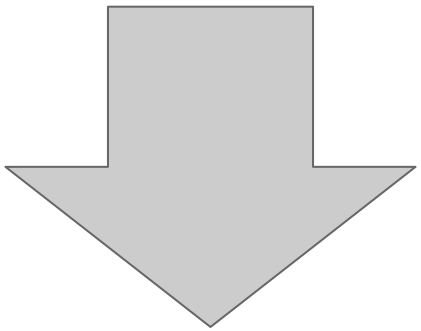
Next Generation Application Platform



*General Purpose Managed Compute Environment

The Real Goal...

```
$ ngap-deploy earthdata-search edsc.tar
```



https://search.ngap.uat.earthdata.nasa.gov/

Application-Hosting & Processing in the Cloud

- Paradigm Shift: How We Implement Systems

  Systems may utilize:

  - Infrastructure as a Service (machine images)
  - Platform as a Service (web servers)
  - Software as a Service (databases, searching)

- Key Benefits

  - Faster time to initial release
  - More effort on software vs. hardware
  - Smaller custom code footprint (outsource non-domain-specific functions)
  - More code and service reuse

Big Questions

1. How can we supply data to all users on a non-discriminatory basis?

Big Questions

2. How can we avoid vendor lock-in for:
   a. Storage
   b. Vendor-unique services

Big Questions

3. How can we attract end users to the cloud, and how will that alter usage patterns?

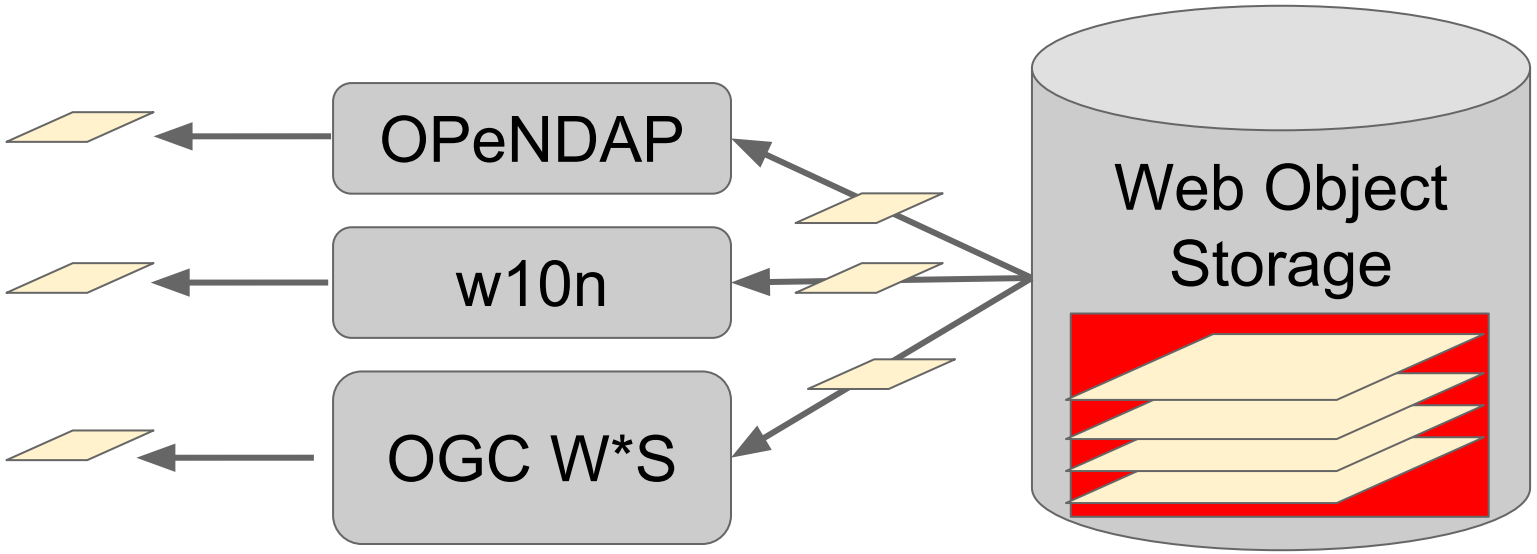
Big Questions

4. How can we predict pricing 2-5 years out...

...especially as usage patterns change

Big Questions

5. How can we migrate near-archive data services to Web Object Storage?



Big Questions

6. What functionality or data should NOT go into the cloud?

Big Questions

7. How do we handle provisioning and accounting of cycles and storage across the DAACs*?

* 4 NASA Centers, 2 Other federal agencies, 4 Universities

Big Questions

8. Do we need new operations policies or procedures?

Big Questions

1. How can we supply data to all users on a non-discriminatory basis?
2. How can we avoid or mitigate vendor lock-in?
3. How can we predict pricing 2-5 years out?
4. How can we attract end users to the cloud?
5. How can we migrate data-proximal services to Web Object Storage?
6. What functionality or data should NOT go into the cloud?
7. How do we handle provisioning and accounting of cycles and storage across the DAACs?
8. Do we need new operations policies or procedures?